Towards a Generative Theory of Hip-Hop[*]

Jonah Katz, MIT Department of Linguistics & Philosophy

1 Introduction

Textsetting, the matching of linguistic objects and rhythmic ones, is a subject of enduring interest to researchers. Most studies of textsetting involve relatively simple forms such as children's songs (Halle 1999), folk verse (Hayes & MacEachern 1996), and art verse similar to these traditions (Lerdahl 2001a).

Hip-hop departs from many of the conventions of these genres, introducing additional levels of complexity. In this paper, I offer an analysis of hip-hop rhyme based on Lerdahl's (2001b) analogy between rhyme and harmony in music. I argue that adopting the formalism of *prolongational reduction* from Lerdahl & Jackendoff's (1983) seminal work *A Generative Theory of Tonal Music* (hencforth GTTM) allows us to analyze and actually predict many of the complex rhythmic phenomena encountered in modern hip-hop.

[*] Thanks to Adam Albright for guidance, discussion, code, and debugging help. Thanks to Fred Lerdahl for sparking my interest in this topic and for being unstintingly generous with his time, work, and criticism. Thanks to David Pesetsky for extensive discussion and suggestions on this and related work. Thanks to Edward Flemming and Donca Steriade for useful discussion. A draft is in progress; if you're interested in talking about the material or seeing the corpus, please email me at jikatz@mit.edu. Audio files for almost all examples can be found at http://web.mit.edu/jikatz/Public/HipHopAudio/.

2 Background

2.1 Hip-Hop

- A musical form that arose in the late 1970s in New York
- Involves setting texts to a rhythm, without having pitches associated with the text
- Like verse poetry and unlike instrumental music: fitting a linguistic consituent to a rhythmic one
- Like (most Western) instrumental music and unlike verse poetry: meter (pattern of beats) is temporally rigid, or *isochronous*:
    - No leeway for stretching or contracting the length of time between beats
    - In this respect, more like jazz or rock music than Western Art ('Classical') Music, which does allow some rhythmic flexibility at some junctures
- Still a matter of debate whether verse poetry involves 'silent beats' or not (see Lerdahl 2004 and Fabb & Halle 2007 for opposing views)
- Isochronous nature of music in general and hip-hop in particular makes it immediately obvious that silent beats do occur in these forms
- Most relevant difference between hip-hop and verse poetry for this paper is that *enjambment*, mismatch between linguistic

and rhythmic constituents, is more common in hip-hop than most kinds of verse poetry

- gives us particularly fertile ground for examining the interacting alignment of musical, linguistic, and rhyme domains

2.2 The Corpus

- I assembled a corpus of 1,097 lines of hip-hop for the purpose of studying rhymes:
  - includes 13 songs by various artists
  - selected mainly for high incidence of complex and/or imperfect rhymes

- Lyrics transcribed in English orthography, then converted to phonetic transcription by script using CMU pronouncing dictionary

- Dictionary transcriptions changed to reflect certain aspects of African American Vernacular English (see Green 2002 for an overview)

- Some stress transcriptions changed to faciliate string alignment by computer program

- 'Lines' identified by listener intuition and coded in corpus as line breaks

- Script identified candidate rhyme domains and cut them down to feasible candidates on the basis of phonetic distance

- Final culling was done by hand to reflect listener intuitions.

- Process resulted in:
  - 1,090 rhyming pairs
  - 736 unambiguous imperfect correspondences between segments
  - 757 corresponding pairs of lines
  - 289 multiple rhyme correspondences (details later)

All statistical generalizations come from the corpus, but specific examples cited later may not.

3 What Rhymes?

Intuitively, a rhyme domain consists of the rime of a stressed syllable and zero or more succeeding unstressed syllables.

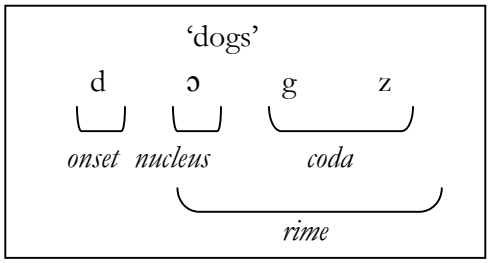
Rhyme domains

Pair	rhyme domains	rhyme?
beat-seat	/it/-/it/	Y
beat-suit	/it/-/ut/	N
barrier-carrier	/æriɹ/-/æriɹ/	Y
barrier-various	/æriɹ/-/æriəs/	N (?)

This shouldn't be controversial, but it's always good to test common sense empirically. In what follows, I test the line endings picked out by (my) listener intuition for various phonetic properties. Of course, the selection of those line endings is itself somewhat subjective; a future study might examine the consistency of these judgements across speakers.

3.1 Stresssed rimes correspond

Rime: string of segments beginning at the nucleus of a syllable and extending to the end.



In the corpus:

- About 94% of stressed syllable pairs singled out as rhyme beginnings have the same vowel.
- Most of remaining cases are relatively 'close' mismatches, e.g. /ɪ/ vs. /ɛ/.
- For about 56%, entire rime is identical in the two strings.
- This means that about 40% of syllable-pairs have different codas.

3.2 Onsets don't correspond

In fact, it looks like they correspond significantly worse than chance:

- 917 onset-pairs from actual rhyme domains contained significantly smaller proportion of perfect matches than sample of 20,000 random onset-pairs from corpus: Fisher's Exact, p=0.002, odds ratio (real/rand) = 0.68
- Phonetic/phonological distance measured using Frisch's shared natural class metric (Frisch et al. 2004) and Albright's alignment algorithm (Albright n.d.). Median distance for the rhyme domain onset-pairs was larger than for the random pairs, 0.87 vs. 0.76.
- Results may indicate that onsets preceding rhyme domains are required to contrast maximally.

3.3 Succeeding unstressed syllables correspond

Examining all sequences of unstressed syllables that follow a stressed syllable in a rhyme domain (including null sequences):

- About 28% of all pairs match perfectly, vs. about 2.5% in a sample of 20,000 random pairs: Fisher's Exact, p<0.001, o.r. = 10.8
- Median Frisch distance per segment (D/S) is smaller in real pairs than random ones, 0.40 vs. 0.58.
- This is actually an overly conservative test, because the random pairs were constrained to match for number of syllables in each string. The real pairs occasionally mismatch for number of syllables.

So this seems relatively straightforward, but just as a sanity check, let's make sure that this isn't just an effect of being *close* to a rhyming stressed syllable.

3.4 Preceding unstressed syllables don't correspond

Examining all sequences of unstressed syllables (including null sequences) that precede the first rhyme domain of a line:

- About 3% of all pairs match perfectly, similar to the random sample.

- Median D/S for these pairs is a whopping 1.2; reflects the fact that strings are not particularly likely to match for number of syllables, leading to messy alignment.

Henceforth, I'll use the notion of 'rhyme domain' from the beginning of section 3:

> A rhyme domain consists of the rime of a stressed syllable and all succeeding unstressed syllables up to the next rhyme domain.

4 Which Rhymes Rhyme Better?

4.1 What is 'better'?

- An idea that's out there: corresponding rhyme domains are required to be phonetically/perceptually similar to one another (Steriade 2003, Kawahara 2007).

- Obviously, the most similar pairs of segments are identical.

If perceptual similarity is driving rhyme, we expect some kinds of mismatch between domains to occur more frequently than others. This is true in the current study:

- Non-parametric Kendall correlation shows that Frisch distance is a significant predictor of observed over expected (O/E) values for mismatches: $\tau = 0.297$, $p < 0.001$.

- O/E values reflect how often mismatches occur in the corpus relative to how often we'd expect them to occur if nothing constrains rhyme correspondence.

- Specific findings about which mismatches are more or less common than expected basically replicate Kawahara's (2008) results for Japanese imperfect puns:
    - Disagreements in place are more common for nasals than voiced obstruents (voiceless obstruents are somewhere in between).
    - Disagreements in voicing are far more common than disagreements in continuancy or nasality.
    - Sonorant sounds and taps (from intervocalic /t/ and /d/) are more likely to correspond with nothing (i.e., insertion/deletion).
    - Voiceless obstruents in English, unlike Japanese, correspond relatively frequently with place mismatches.
    - Unlike Japanese, coronal obstruents are relatively likely to correspond to nothing; this may be due to their appearance at the ends of consonant clusters, where they may be weakened or absent phonetically.

- See Kawahara (2007, 2008) for arguments that many of these effects are predicted by independently-motivated generalizations about perceptual similarity, and are not predicted by measures of phonological similarity (i.e., distinctive features).

- Phonetically more similar rhyme domains are better rhymes. I'll use D/S as a rough measure of phonetic similarity.
    - It's an approximation though, missing some generalizations, and excluding context effects (i.e., /t/ and /d/ are more similar intervocallically than they are word-finally).
    - D/S also produces *non-parametric* data; values are not normally distributed and it's probably not the case that D/S is an interval measure.

- Because hip-hop is grounded in isochronous meter, mismatching numbers of syllables in corresponding rhyme domains should be especially marked. I'll take syllabic matching as another indicator of rhyme goodness.

- As we've seen, stressed vowels have a special salience in rhyme, marking the beginning of domains and almost always matching in corresponding domains. I'll take stressed vowel match as an indicator of rhyme goodness.

- This leaves us with four measures of rhyme goodness:
    1. Probability of perfect match
    2. Probability of syllabic match
    3. Probability of stressed vowel match
    4. Frisch D/S (tentative)

4.2 The Last Effect

A rhyme only needs to have a single corresponding rhyme domain. This is the 'usual' way we think of rhyme (rhyme domains are underlined henceforth):

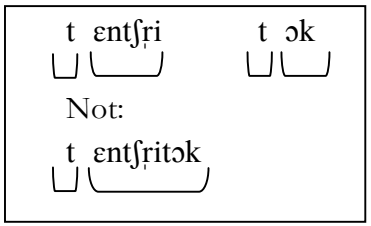
One day when I was chillin' in Kentucky Fried Chicken
Just mindin' my business, eatin' food and finger lickin'
-Run DMC, *You be Illin'* (1986)

In modern hip-hop, artists often stack rhyme domains at the right edge of constituents. These are known as *multis*:

2-domain multi:
with penitentiary talk
coke and a hennessy walk
-Common, *Cold Blooded* (2000)

3 (or 4?) -domain multi:
at the real estate, behavin' type choosy
want a palace with the shit beige and light blue please
Slick Rick, *Street Talkin'* (1999)

Note that these are not simply cases of long rhyme domains. Each domain is independent, with a contrasting onset followed by a matching stressed rime and string of unstressed syllables:



I'll refer to a series of stacked rhyme domains as an *R-complex*, indicated here by brackets:

at the real estate, be[havin' type choosy]

We can now ask questions about rhyme domains occupying different positions in an R-complex. One intuition, which sparked this whole study in the first place, is that rhyme correspondence in multis is 'tighter' the further one looks toward the end of the R-complex. A key finding from the corpus study:

Rhyme domains that are last in their R-complex correspond more closely than those that are non-last.

This is indicated by the four measures discussed above:

- Last domains have a higher probability of corresponding perfectly than non-last domains, 44% vs. 29%: o.r. = 1.99, Fisher's exact, $p < 0.001$.
- Last domains have a higher probability of matching numbers of syllables, 99.6% vs. 94.3%: o.r. = 15.17, $p < 0.001$.
- Last domains have a (near-significantly) higher probability of matching stressed vowels, 94.5% vs. 92.2%: o.r. = 1.56, $p = 0.061$.
- Last domains have a lower median D/S, 0.28 vs. 0.35.

4.3 The First Effect

A second finding (this one unexpected):

Rhyme domains that are preceded in their R-complex correspond more closely than those that are not preceded (i.e., first).

This seems to be a different kind of effect from the Last Effect. It affects the probability of perfect match and median Frisch D/S, but not syllable or stressed vowel mismatch.

- First domains have a lower probability of corresponding perfectly than non-first, 53% vs. 34%: o.r = 0.46, $p < 0.001$.
- First domains have a higher median D/S, 0.32 vs. 0.27.
- The probabilities of syllable matching and stressed vowel matching don't differ between first and non-first domains; about 94% of stressed vowels match and 98-99% of domains have the same number of syllables.

Summarizing:

- last domains differ from non-last
    - categorical absence of syllable mismatches
    - higher probability of perfect matching
    - higher probability of stressed vowel matching
    - lower D/S
- preceded domains differ from non-preceded
    - higher probability of perfect matching
    - lower D/S

This is not simply a consequence of the fact that first domains are more likely to be non-last. When both factors are considered in the same model of perfect rhyme probability, it becomes clear that there are two independent effects at work.

4.4 Investigating the two effects

- Forward stepwise logistic regression shows that including 'last' and 'first' as predictors of perfect rhyming significantly improves the fit of the model at each step. But adding the interaction term between first and last does not significantly improve fit.
    - For first: B = 0.659, s.e. = 0.138, $p < 0.001$. Exp(B) = 1.932.
    - For last: B = -0.528, s.e. = 0.146, $p < 0.001$. Exp(B) = 0.590.

- Hierarchical loglinear analysis with backward elimination confirms this result. Starting with the saturated model (simplifying a bit, this is just the data itself), we find that removing the interaction between first, last, and rhyme perfection does not significantly affect the fit of the model. But removing either interaction between rhyme perfection and first/last significantly decreases the fit of the model.
    - For first x perfect, change in $\chi^2 = 22.8$, $p < 0.001$.
    - For last x perfect, change in $\chi^2 = 13.3$, $p < 0.001$.
    - For the final model (generated by two-way terms), $\chi^2(1) = 0.612$, $p = 0.43$.

- So we have two separate, independent effects:
    - Last rhyme domains correspond in a qualitatively different and closer way.
    - Preceded rhyme domains correspond in a gradiently different and closer way.

5 Basics of Textsetting[1]

5.1 Consituents

Verse fundamentally involves putting together a linguistic consituent and a rhythmic consituent.

- The linguistic consituents are generally hypothesized to be *prosodic phrases* (Selkirk 1986, Nespor & Vogel 1986 for prosodic phrases).

- The rhythmic constituents in verse poetry are often hypothesized to be *lines*, and various sub- and super-line levels (e.g. hemistich, couplet). Hayes & MacEachern (1996), for instance, define lines in folk verse/song by their propensity to coincide with prosodic constituents and their tendency to have a pause at the end.

- Generally, a line has a certain number of strong and weak beats, or feet.
    - e.g., iambic pentameter is five weak beats interspersed with five strong beats; equivalent to five iambic (weak-strong) feet.
    - Quatrain form is defined by having four beats with weak subdivisions, but the subdivisions don't always coincide with linguistic events (Hayes & MacEachern 1996, Lerdahl 2004)

[1] The basics of textsetting are discussed in Halle & Lerdahl (1993), Halle (1999 *et seq.*), Hayes (1989 *et seq.*), Lerdahl (2004). Generalizations about 'the literature' are based on these sources.

5.2 Alignment and enjambment

Most folk verse, children's songs, verse poetry, etc. show a strong tendency to have lines and prosodic constituents align exactly:

(shall	Í)	(com-	páre	thee)	(to	a	súm-	mer's	dáy)
·	_	·	_	·	_	·	_	·	_
(thou	art	more	lóve-	ly)	(and	more	tém-	pe	-rate)
·	_	·	_	·	_	·	_	·	_
(róugh	wínds)	(do	sháke)	(the	dár-	ling	búds	of	Máy)
·	_	·	_	·	_	·	_	·	_
(and	súm-	mer's	léase)	(hath	all	too	shórt	a	dáte)
·	_	·	_	·	_	·	_	·	_

Shakespeare, Sonnet 18. "." = weak beat; "_" = strong beat; ( ) = plausible prosodic phrase

Departures from this canonical alignment produce verse objects that are distinctly marked (Halle 2003). Consider the following slightly altered version of the Shakespeare sonnet, where stress still conforms to iambic pentameter but alignment is violated:

Shakespeare vandalized.

(shall	Í)	(com-	páre	thee)	(to	a	súm-	mer	báll-
·	_	·	_	·	_	·	_	·	_
game)	(with	some	lóve-	ly	pláy-	ers)	(rún-	ning	thróugh
·	_	·	_	·	_	·	_	·	_
the	wínd)	(which	shákes)	(the	dár-	ling	búds	of	Máy)
·	_	·	_	·	_	·	_	·	_
(and	súm-	mer's	léase)	(hath	all	too	shórt	a	dáte)
·	_	·	_	·	_	·	_	·	_

- Though marked, this structure does occur in verse poetry. It is known as *enjambment*.

- Here are some cases of enjambment in verse poetry:

That's my last Duchess painted on the wall,
Looking as if she were alive. I call
That piece a wonder, now....

Browning, *My Last Duchess*

A thing of beauty is a joy forever:
Its loveliness increases; it will never
Pass into nothingness but still will keep
A bower quiet for us, and asleep
Full of sweet dreams, and health, and quiet breathing.

Keats, *Endymion.* (ll .1-5)

- In each of these cases, the end of a rhythmic unit (a line) fails to coincide with the end of a linguistic unit (a clause or intonational phrase).

- One important observation here:

When enjambment occurs in rhyming verse, the rhyme relation always holds between the final domains in the rhythmic line, *not* the final domains in the linguistic phrase.

- This means that, if we were to make a statement about what rhymes in verse, it should take this form:

The rhyme relation must hold between the final domain of a rhythmic group (line) and the final domain of the group that follows it within the couplet.

6 Why is Hip-Hop (Structurally) Interesting?

6.1 Shakespearean hip-hop

- Like verse poetry, early (roughly 1980s) hip-hop is characterized by mostly perfect textsetting:
    - Lines of equal rhythmic duration and sentences align with one another
    - Rhyme relation must hold between final domain of a line and and final domain of following line.

Here's an example:

One day when I was chillin' in Kentucky Fried Chicken)
Just mindin' my business, eatin' food and finger lickin')
This dude walked in, lookin' strange and kind of funny)
Went up to the front with a menu and his money)

Run DMC, *You be Illin'* (1986).

- In this figure and henceforth:
    - line breaks indicate moments of temporal disjuncture in the musical surface. That is, if you wrote out this sequence in musical notation, the line breaks would correspond to rests or long notes.
    - The right parentheses indicate the end of linguistic consituents (these could be identified as *intonational phrases* in the current example).
    - The underlines indicate rhyme domains.

6.2 Keatsian hip-hop

- As in verse poetry, enjambment became fairly standard in hip-hop as the form progressed.

- Below we see some cases of enjambment in hip-hop. As in verse poetry, the rhyme correspondence here holds across line-final elements:

…by these dusty blunted
cats) who rap like they don't know)
that the fact is that they being hunted)

The Roots, *Respond/React* (1996).

I'ma go for mine)
You know the time)
Now that I'm older) I'm
gonna…

Del, *Catch a Bad One* (1993).

And you know)
prob'ly get cussed) if I
backslap miss) while she's busy tryin' to justify)

Slick Rick, *The Art of Story Tellin'* (1999)

6.3 e.e. cummings-ian hip-hop?

- As seen above, rhyme often 'follows' rhythmic grouping in cases of enjambment in hip-hop.

- This is entirely identical to the situation in verse poetry, so we can still hope that rhyme will be an orderly right-edge phenomenon characterized by this statement from section 5:

The rhyme relation must hold between the final domain of a rhythmic group (line) and the final domain of the group that follows it within the couplet.

- But hip-hop is not so simple.

- In some cases of enjambment, the rhyme relation holds between the right edges of *linguistic* constituents instead:

…is in the pocket like wallets)
I got the bounce like hydraulics)
they can't call it)
I got to swerve like alcohol)
ics) my freshman year…

Kanye West, *Get 'em High* (2004)

ginseng tree trunks)
rockin' the p
funk) rockin' her knees
up) champion lover not ease up)

Mos Def, *Ms. Fat Booty* (1999)

out the frying pan)
into the fire) trans-
form into the ghost rider)

Method Man, *Triumph* (1999)

- But in many cases, the rhyme domain isn't at the edge of a linguistic *or* a rhythmic constituent:

But you don't really hear me) though I
spit it(clearly) so it's live
out your stereo)

Talib Kweli, *Listen* (2007)

So now I'm jettin' to the building lobby)
And it was full of children) prob'ly couldn't see as high as I be)

Nas, *New York State of Mind* (1994)

- In fact, rhyme domains are often scattered throughout the musical surface, corresponding to no particular grouping, rhythmic or otherwise:

I bomb atomically)
Socrates' philosophies and hypotheses)
can't define) how I be droppin' these
mockeries)
lyrically perform armed robbery) flee with the lottery)
possibly) they spotted me)

Inspectah Deck, *Triumph* (1999)

- This suggests that any attempt to characterize rhyme as a relation holding between edges of some pre-defined domain is doomed.

- These examples, which are not canonical but are certainly not rare, show that rhyme is partially independent from rhythmic factors.

- Rhyme, then, is going to need its own level of representation in any model of hip-hop.

- In the next section, I'll suggest that music theory offers us an excellent way of representing rhyme relations, developing an analysis introduced by Lerdahl (2004).

7 GTTM to the rescue

It appears that long-distance thematic (rhyme) relations may hold between elements that do not occupy parallel rhythmic positions. Put another way, it is impossible to state the conditions on what rhymes with what if we only allow ourselves to make reference to linguistic, musical, and/or metrical constituents.

Music theory offers an instructive parallel here: Western tonal harmony is also a system where long-distance thematic (tonal) relations often cross-cut rhythmic boundaries.

In GTTM, this situation is explained by positing separate components for *time-span* and *prolongational reduction*. The time-span reduction captures patterns of rhythmic and structural prominence; the prolongational reduction captures thematic

relations between elements, using structural prominence as a starting-off point.

In section 7.1, I'll attempt a brief introduction to reductions as models of music. This is a complicated topic that is worth pursuing in more detail if you have any interest in music; the most comprehensive statement of the system comprises the middle chapters of GTTM.

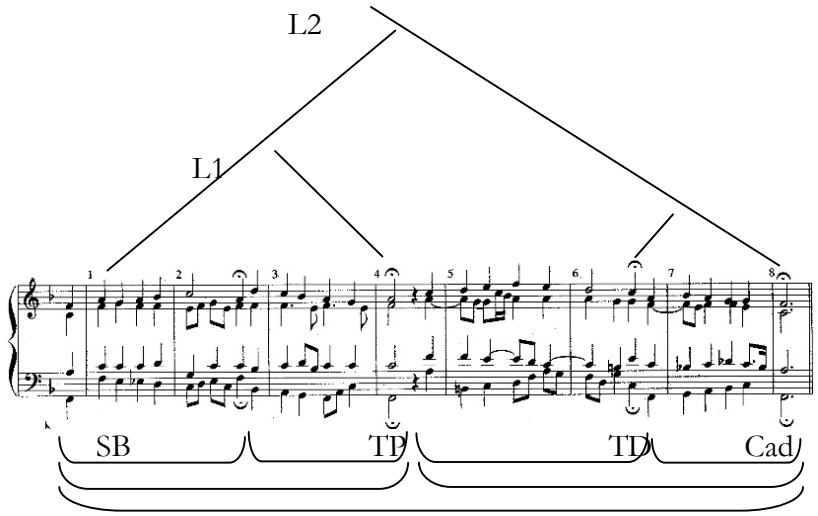
7.1 The derivational system[2]

7.1.1 TIME-SPAN REDUCTION

- Operates recursively, starting at the level of the smallest rhythmic groups in the piece.
- Picks out the most stable or structurally important event in each rhythmic group, and subordinates all other events to it.
- Most important event is carried up to next level as head prominence, and becomes an input to the next iteration.
- System is formally identical to headed prosodic trees from linguistic theory.

[2] The analyses in this section are taken from Lerdahl (2001b), Ch. 1, and Lerdahl & Jackendoff (1983), Ch. 8. My exposition of the principles at issue is somewhat different from those discussions, which are far more extensive.

Take as an example a simplified version of Lerdahl's (2001b) analysis of the Bach Chorale 'Christus, der ist mein Leben' (BWV 281):



- Abbreviations below are for descriptions of harmonic function:
  - SB = structural beginning
  - TP = tonic prolongation
  - TD = tonicized dominant
  - Cad = Cadence
- Bracketing below indicates rhythmic constituency:
  - At the lowest level, we have four groups of two measures, each of which has a head (the harmonic events listed above).

- At level 1, the piece breaks down into two four-measure groups; SB is the head of the 1st and the cadence is the head of the 2nd.
- At the highest level (L2), the piece forms a coherent whole, with the cadence as its head.

- Time-span analysis captures several insights about the piece:
  - Consists of four temporally-regular two-bar constituents
  - These units group into two binary four-bar groups, punctuated by long rest in m. 4.
  - Each group culminates in a head: the head is the tonic of the larger piece's tonal area (F major) in all groups except the third, where it is the tonic of C major (V of home key).
  - This unit is basically just like a prosodic phrase, with hierarchical sub-grouping and headedness.

- Time-span analysis doesn't tell us anything about the tonal ebb and flow of the piece, patterns of relative tension and relaxation between events and groups.

- In this example, I've given some indication of harmonic function with abbreviations below the score. But deriving these functions is the job of prolongational analysis.

7.1.2 PROLONGATIONAL REDUCTION

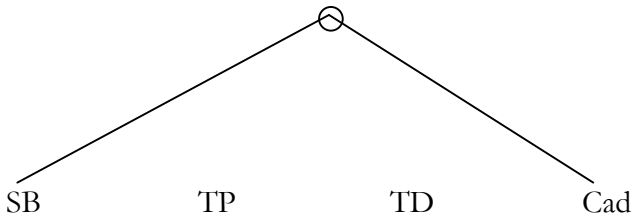
- Begins with the most rhythmically prominent events in the time-span reduction: this is L2 in the current example.

- Branchings don't represent relative structural prominence; they encode thematic relations between elements:
  - *Prolongation* is repetition, and it is the most stable way for two events to relate to one another.
  - *Progression* is some non-null movement from one event to another, and its stability depends on a distance metric (developed in Lerdahl 2001b).

- Each level of prolongational analysis makes the most stable attachments possible of events from the most prominent level of time-span reduction.

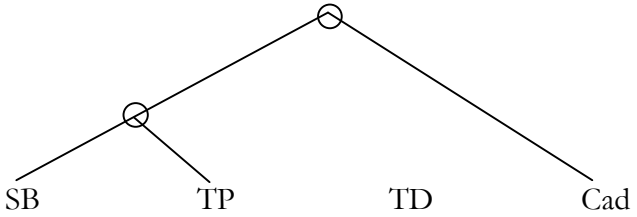
- For the current example, prolongational analysis begins with the two most rhythmically prominent events, SB and Cad.

- They form a strong prolongation, indicated with a circle.



- At the next-highest level of time-span prominence, TP and TD become accessible to the prolongational system.

- TP forms the most stable possible connection with SB (prolongation), so it is attached there:



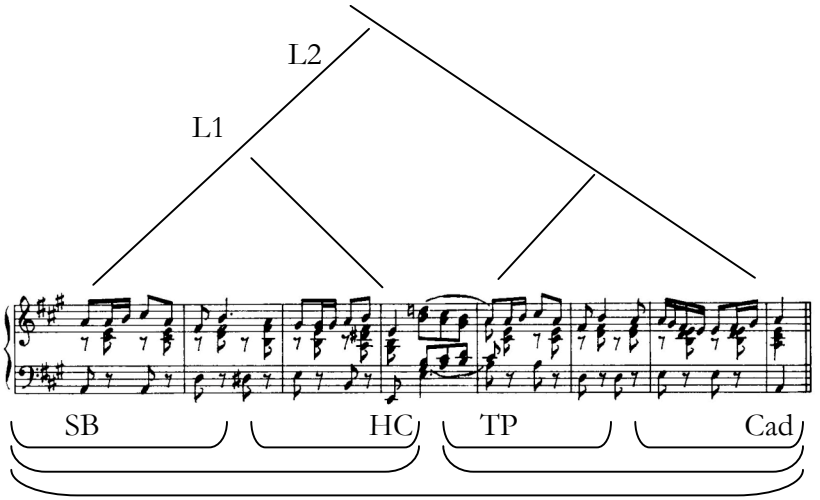
- Analysis proceeds in this manner until the whole piece has been assigned structure.

7.2 Prolongation and rhythmic independence

7.2.1 EDGE MISMATCHES

- For the example above, prolongational analysis didn't change the constituency from time-span analysis.
    - [SB-TP] is a rhythmic consituent (time-span), and also a thematic consituent (prolongation).
    - So we could have just done the prolongational analysis by looking inside time-span consituents and stating the thematic relations that hold between the heads of those constituents.
    - Just like in hip-hop, however, trying to state thematic relations over rhythmic domains is only possible for the simplest pieces.

- Consider now the opening of the aria "La ci darem la mano" from Mozart's Don Giovanni (K.527):



- The new abbreviation here 'HC' stands for half-cadence.

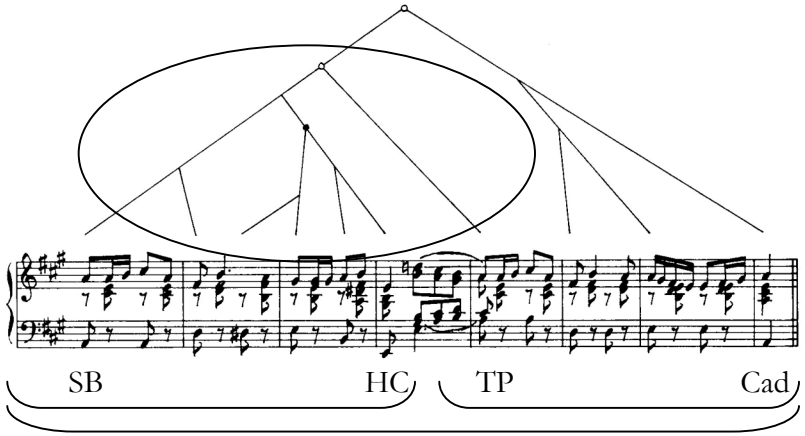
- This excerpt differs in harmonic rhythm from the Bach excerpt:
    - It is an antecedent-consequent period, a very frequent type of phrase in Classical-era art music.
    - The antecedent is four bars, further subdivided into two two-bar groups.
    - The consequent is parallel to the antecedent, but ends in tonal closure (cadence).
    - The most important feature of this phrase-type for now is that the tonic prolongation occurs at the beginning of the consequent, that is, in the second half of the phrase.

- Although the harmonic *rhythm* of the two excerpts differs, they have a number of harmonic properties in common:

- Both begin with a tonic chord and end with a cadence on that same chord.
- Both involve a departure from tonic (not analyzed in the Bach example) followed by a return to that tonic (the TP event).
- The TP event is what we'll focus on here.

- As before, the two most prominent events in the excerpt form the highest level of the prolongational reduction, and lower events attach in order of rhythmic prominence and thematic stability.

- The final prolongational analysis from GTTM (p. 200), is shown below:



- Note that this analysis does not preserve time-span consituency: [[SB-HC]-TP] is not a rhythmic consituent, but it is a thematic constituent.

- This reflects the fact that thematic relations are partially independent from rhythm.

- The relationship between a tonic and its prolongation is the *same relationship* regardless of whether the prolongation occurs in the 4th measure of a phrase or the 5th.

- This type of branching mismatch is how prolongational analysis allows us to represent thematic relations that cut across rhythmic constituents.

- Prolongational analysis also allows thematic relations to be partially independent from rhythmic *prominence*, through the *Interaction Principle*.

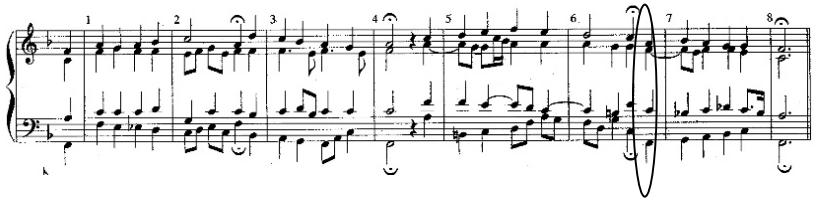
7.2.2 PROMINENCE MISMATCHES

- As introduced above, prolongational analysis enforces strict correspondence between rhythmic (time-span) prominence and thematic (prolongational) prominence.

- Every level of time-span relations is mapped onto a level of prolongational relations.

- This strict condition does not adequately capture our intuitions about tonal harmony.

- Returning to the Bach chorale, we focus now on the tonic return at the end of m. 6 (circled):

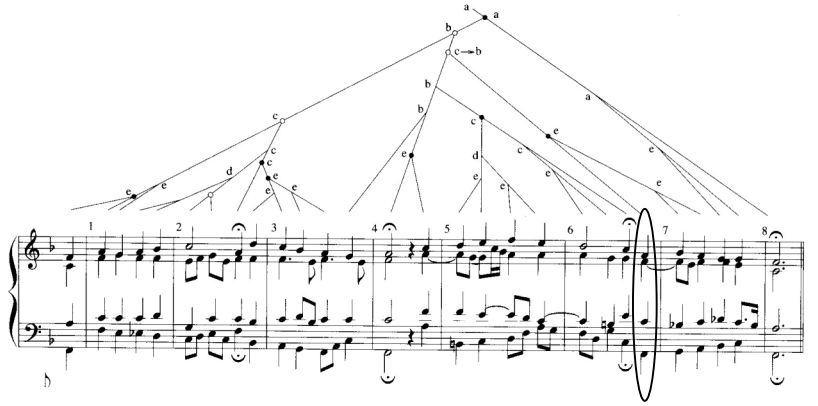


- Rhythmically, this event is overshadowed by the cadence at the end of the piece.

- The tonic return is not accessible at the two-bar level of the time-span reduction (see diagram p. 14, sec. 7.1).

- But this event is very important thematically; it marks the return of the tonic after departing to the key of the dominant in the third two-bar group.

- This is reflected in the final prolongational analysis of the piece from Lerdahl (2001b, pg. 22):



- In this analysis, the tonic return in m.6 is rightly seen as a prolongation of the tonic in m.4.

- This type of mismatch between rhythmic and thematic prominence is licensed by the *interaction principle*:

Interaction Principle: Prolongational analysis may attach an event at a level higher than its time-span prominence if it results in a more stable prolongational connection.
(simplified from GTTM, Lerdahl 2001b)

- How to measure prolongational stability is addressed in magnificent and punishing detail in Lerdahl (2001b).

- The only part that need concern us here is that prolongation is more stable than progression.

- Therefore, analyzing the tonic return as a prolongation of a distant event results in a more stable connection that analyzing it as a progression to or from a more local rhythmically prominent event.

- This is a second kind of flexibility between rhythmic and thematic relations that GTTM captures.

Summarizing:

Prolongational analysis allows us to describe how thematic dependencies may cross-cut rhythmic/structural dependencies, and allows us to describe how events that are relatively weak from a rhythmic standpoint may be relatively prominent from a thematic standpoint.

These are exactly the problems we ran into with hip-hop.

8 Of Frost and Wu-Tang

8.1 Rhyme as prolongation

- The analytical system for rhyme introduced in Lerdahl (2004) is rooted in an analogy between rhyme and prolongation:
    - Both are a kind of repetition.
    - Both work to thematically join elements from the beginnings and ends of larger pieces.
    - I'll argue here that, at least for hip-hop, the independence of rhyme from rhythm is formally identical to the independence of musical prolongation from rhythm.

- Lerdahl's approach is illustrated with the Robert Frost poem 'Nothing Gold Can Stay':
    - The poem is in iambic tetrameter, with a silent beat at the end of each line.
    - Each line corresponds to a linguistic intonational phrase.
    - Rhymes are located at the right edge of every line, and are monosyllabic (with one possible exception).

- The rhythmic analysis of verse consists of separating it into prosodic constituents, linguistic units that are phrased together.

- Like time-spans in music, prosodic constituents have heads. These are the most highly stressed elements in each prosodic domain.

- This system allows an elegant account of canonical verse, like the Frost poem or the early hip-hop described in section II:
    - Linguistic prosodic phrases (in English) have their highest stress (rhythmic) prominence on the right: this is known as the Nuclear Stress Rule (Chomsky & Halle 1968).
    - This means that at the couplet level, the two most prominent elements will be the final stress of each line.
    - The prolongational component then joins these two prominent elements together as a prolongation (rhyme).
    - And that's why rhymes occur at right edges of lines.
    - Here's an example from Lerdahl's paper, with headed brackets in place of a time-span tree:

1-x level:	[Na-][ture's]	[first]	[green]	[is]	[gold]	[Her	hard-][est]	[hue]	[to]	[hold]
2-x level:	[Na-]	[first]	[green]	[	gold]	[	hard-]	[hue]	[	hold]
3-x level:	[Na-]	[	green]	[	gold]	[		hue]	[	hold]
4-x level:	[				gold]	[				hold]

- At the beginning of prolongational analysis, the most prominent elements, at the 4-x level, will be linked as a rhyme.

8.2 Added complexity I: cross-branching

- If rhyme is really a kind of prolongation, we predict several types of rhythm-rhyme dissociation that are not observed in Frost.

- First, Frost's rhymes always involve two prominent elements within one couplet; but because of the flexibility in prolongational branching, we should also see branchings cutting across couplets.

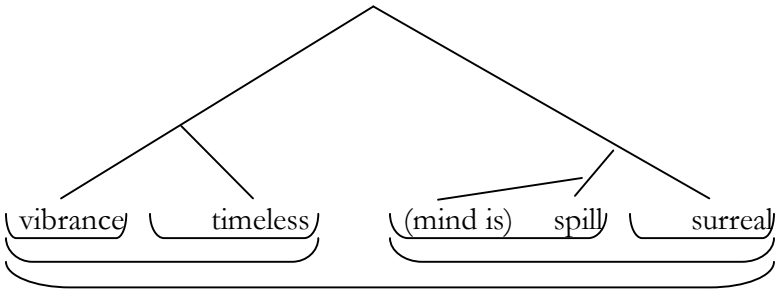
- Hip-hop confirms this prediction:

stylistic violence with vibrance
the sign o' times with rhyme shit that's timeless
a mind is a terrible thing to spill
rap life's like a dream it seems surreal

Common, *Cold Blooded* (2000)

- There are two couplets in this excerpt, with each line marked by linguistic constituency, temporal pause, and roughly equal rhythmic duration (8 beats).

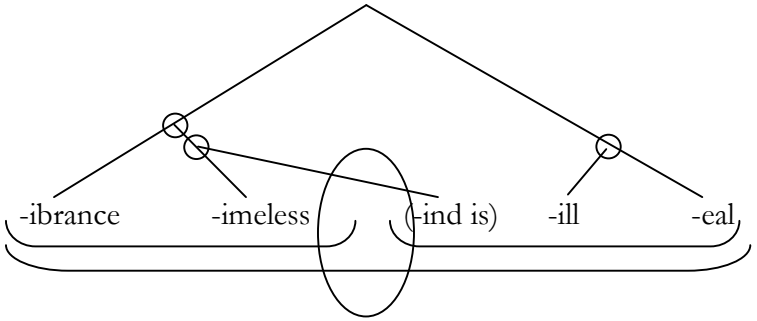
- Here's a time-span reduction to the two-measure level, with the less-prominent *mind is* in parentheses:



- Rhythmically, *mind is* is part of the second couplet.

- But it forms a stable prolongational connection (rhyme) with elements in the first couplet.

- This results in a prolongational structure that cuts across the rhythmic grouping of the piece:



8.3 Added complexity II: the Interaction Principle

- The prolongational analysis of rhyme also predicts that the rhyme relation doesn't necessarily need to hold between the two most prominent elements in a rhythmic domain.

- Hip-hop also confirms this prediction:

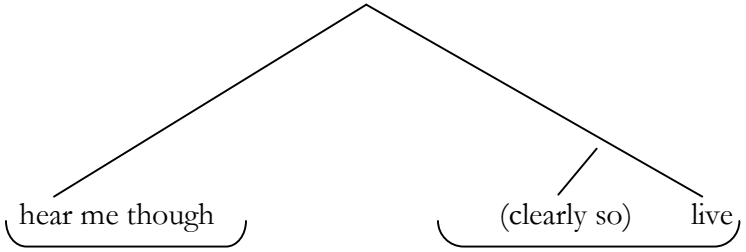
But you don't really hear me) though I
spit it clearly) so it's live
out your stereo)

Talib Kweli, *Listen* (2007)

- It's unclear in this example what the most rhythmically prominent element is in the first line (enjambment makes it hard to tell).

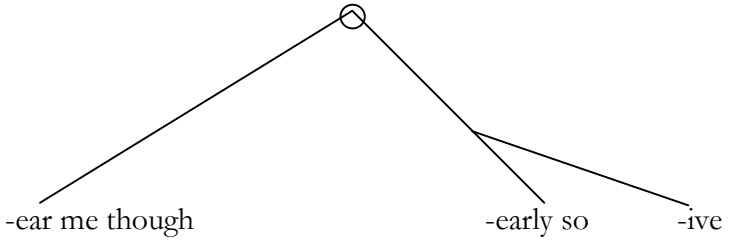
- But the most prominent element in the second line is clearly *live*:
    - rightmost stress in the group
    - falls on a metrical downbeat
    - longer duration than either syllable flanking it

This means that the time-span representation at the level of the line is, loosely:



- At the first step of prolongational reduction, *clearly so* forms a more stable prolongational with the preceding material than the more rhythmically prominent *live* does.

- The Interaction Principle is invoked, resulting in 'promotion' of the rhyming material past the time-span head:



8.4 Added complexity III: inners

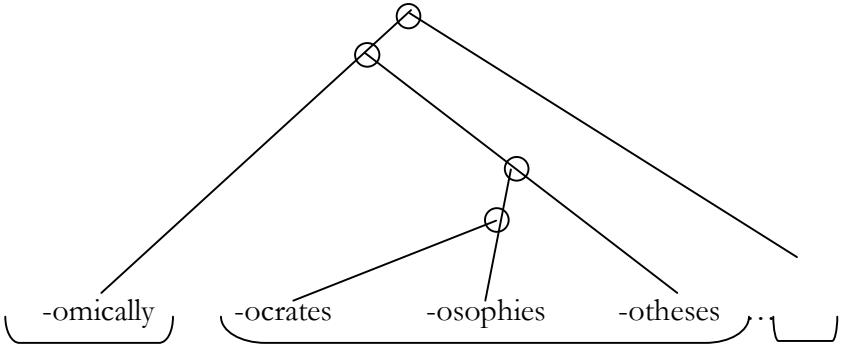
- Prolongational relations hold between events at all levels.

- Rhyme, therefore, shouldn't be limited to the rightmost or structurally most prominent element in a line; it should occur freely internal to the line as well.

- Hip-hop confirms this prediction:

> I bomb atomically)
> Socrates' philosophies and hypotheses)
> can't define) how I be droppin' these
> mockeries)
> lyrically perform armed robbery) flee with the lottery)
> possibly) they spotted me)

Inspectah Deck, *Triumph* (1999)

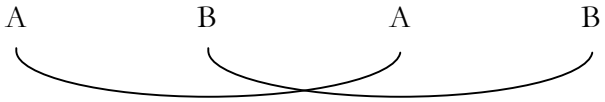


9 Cross-Serial Dependencies

9.1 R-incorporation

- If the phonetic rhyme relation is triggered only when two elements are joined at a node in a prolongational tree, it leaves us unable to explain certain structures.

- These include any situation where more than one rhyme is 'active' at the same prolongational level:
    - Alternating rhymes (ABAB)
    - Stacked rhyme domains (multis)
    - Intermittent rhymes (e.g. ABBAABA)



- This type of relation can be described as parallelism across branches, which I will call R-*incorporation*.

- I propose that R-incorporation is a primitive prolongational relation for rhyme space, more stable than progression.

R-incorporation: For two events e_i e_j such that Rhyme(e_i,e_j), event e_k forms an R-incorporation with e_i iff there exists some e_l such that:

i. Rhyme(e_k,e_l) and

ii. The attachments of e_k to e_i and of e_l to e_j may be analyzed as occupying parallel positions in prolongational structure.

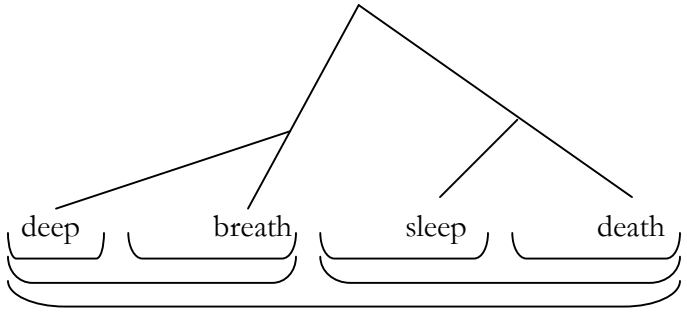
- The process of R-incorporation creates an R-complex, a series of stacked rhyme domains.

Let's examine a simple ABAB rhyme:

inhale deep
like the words of my breath
I never sleep
cause sleep is the cousin of death

Nas, *New York State of Mind* (1994)

Here's the time-span reduction at the line level:

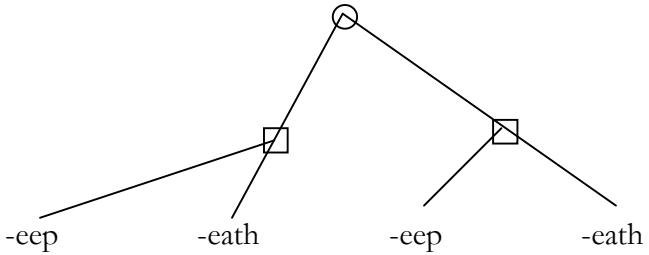


- At the first step of prolongational analysis, *breath* and *death* are attached as a prolongation.

- At the next step, *deep* and *sleep* would normally just form progressions into their non-rhyming neighbors.

- But the definition above allows them to incorporate into the R-complex (indicated below with a rectangle).

- Because R-incorporation is a more stable prolongational relation than progression, this option is preferred:



- In this example, R-incorporation happens at the level of the line (rhythmic group):
    - Because the most prominent element of a line tends to be the last stress, R-incorporation at this level tends to join together the ends of lines.
    - There is other, less rhythmically prominent material intervening between the rhyme domains in the R-complex.
    - This is canonical ABAB stanza form.

9.2 R-incorporation and rhythmic independence

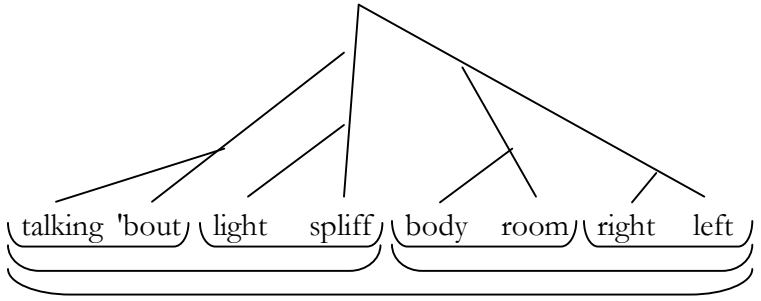
- R-incorporation, being a primitive prolongational relation, should also occur line-internally, and needn't pick out the most prominent rhythmic constituent:
    - The Interaction Principle predicts that incorporated domains can appear in rhythmically weak positions.
    - R-incorporation at this level could result in, for instance, a series of stacked rhyme domains at the right edge of the group.
    - This is a multi, as introduced in Section II.

Here's an example:

talking 'bout light the spliff
everybody in the room swing right to left

Beatnuts, *Prendelo* (2001)

- Here's a plausible time-span representation (the precise relative stress is not crucial):

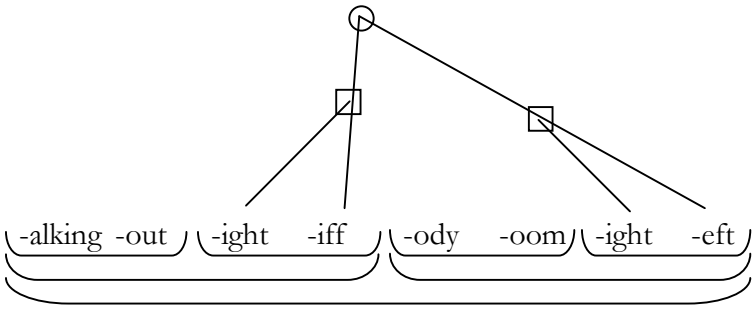


- The first step in prolongational analysis joins *spliff* and *left* together as a prolongation.

- The next-most-prominent elements are *'bout* and *room*.

- But *light* and *right* meet the conditions for R-incorporation, which is a more stable prolongational connection.

- They are promoted past the more rhythmically-prominent elements and attached next:



- This is a canonical multi, with a 2-domain R-complex.

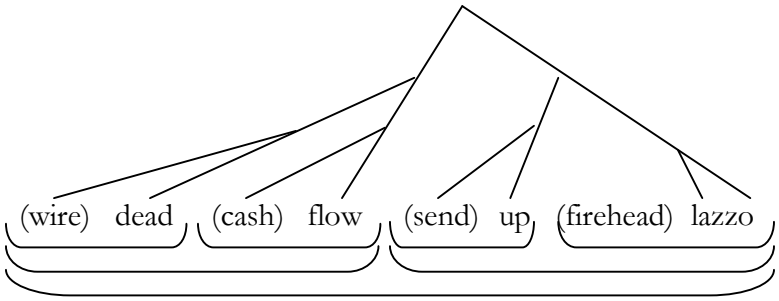
- The framework also predicts that we should be able to mix R-incorporation at different levels of rhythmic prominence.

- This prediction is borne out:

our wire is dead he's in it for the cash flow like to send a big up to Firehead Lazzo

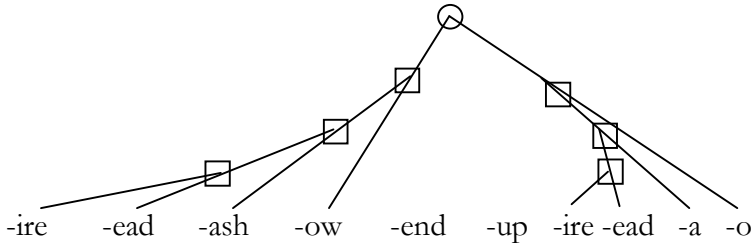
MF Doom, *El Chupa Nibre* (2005)

Time-span reduction:



- Prolongational analysis generates a string of R-incorporations at various levels.

- This results in an R-complex of domains that are string-adjacent in the second consituent, but not in the first:



- This structure is extremely complex or marked; it involves several applications of the Interaction Principle, resulting in a distorted mapping from rhythmic to thematic structure.

10 Summing Up

10.1 The prolongational link

Hip-hop presents several problems for generative textsetting models, displaying a level of dissociation between rhyme and rhythm not seen in other genres.

I've argued here that this dissociation is predicted by a prolongational approach to rhyme, as outlined in Lerdahl (2004). With a few revisions and additions, notably the notion of R-incorporation, all of the hip-hop data can be modeled with prolongational analysis; in fact, the analysis predicts that these structures should exist.

This suggests that the incorporation of thematic properties (pitch-space distance, phonetic rhyme distance) into rhythmic structures can be added to the list of formal similarities between language and music.

10.2 Hip-hop and innovation

Hip-hop rhyme and rhythm possess a level of complexity unseen in folk verse, children's song, and other genres discussed in the textsetting literature.

One difference concerns enjambment; it is rare and marked in the genres mentioned above, but rather frequent in hip-hop.

A second difference is that hip-hop seems to lack a constraint preferring preservation of consituency from rhythmic to thematic analyses, as Lerdahl proposes for the Frost poem.

Finally, hip-hop appears to allow relatively free application of the Interaction Principle, obscuring the relationship between rhythmic and thematic prominence.

Early hip-hop lacks all of these properties, making it essentially like children's song or folk verse.

I speculate that these developments are choices made by performers in order to complicate and diversify the rather basic structures they began with, entirely parallel to developments in art music and art verse.

In art music, composers first weakened (e.g. Wagner) then changed (e.g. Schoenberg) or eliminated (e.g. Cage) constraints on what constitute well-formed harmonic relations.

In art verse, poets weakened and eliminated constraints on the alignment of linguistic and rhythmic objects (resulting in enjambment) and the generation of well formed metrical objects (resulting in the gradual elimination of meter).

In hip-hop, performers have weakened constraints on the alignment of linguistic and rhythmic objects, as in art verse. They also appear to have weakened constraints governing the transparency of the mapping from rhythm to thematic structure, making the thematic (prolongational) component of hip-hop more similar to tonal music than to other genres of textsetting.

Note that these developments complement each other: as the mapping of linguistic material onto musical rhythms gets more irregular and complex, it becomes harder to perceive regular grouping units, such as lines or couplets. Allowing rhymes to

appear stacked at the end of a group (rather than restricting them to rhythmic prominences) should make it easier to segment the musical surface into groups, all else being equal.

References

Chomsky, N. & M. Halle. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Fabb, N. & M. Halle. The grid. Keynote address at Language and Music as Cognitive Systems, Cambridge University, May 2007.

Frisch, S. , J. Pierrehumbert, & M. Broe. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* **22**, 179-228.

Green, L. (2002). *African-American English: A Linguistic Introduction*. New York: Cambridge University Press.

Halle, J. (1999). *A Grammar of Improvised Textsetting*. DMA dissertation, Columbia University.

Halle, J. (2003). Constituency Matching in Metrical Texts. Submitted, proceedings of *Words and Music*, U. of Missouri-Columbia, March 14 2003.

Halle, J., & F. Lerdahl (1993). A Generative Text-Setting Model. *Current Musicology* **55**, 3-26.

Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky & G. Youmans (eds.), *Phonetics and Phonology I: Rhythm and Meter*. Orlando, Fla.: Academic Press. 201-260.

Hayes, B. & M. MacEachern. (1996). Are there lines in folk poetry? In *UCLA Working Papers in Phonology* **1**. 125-142.

Kawahara, S. (2007). Half-rhymes in Japanese rap lyrics and knowledge of similarity. *Journal of East Asian Linguistics* **16**, 113-144.

Kawahara, S. (2008). The role of psychoacoustic similarity in Japanese puns: a corpus study. To appear in *Journal of Linguistics*.

Lerdahl, F. (2001a). The Sounds of Poetry Viewed as Music. In R. Zatorre and I. Peretz, eds., *The Biological Foundations of Music*. Annals of the New York Academy of Sciences, 930, 337-354.

Lerdahl, F. (2001b). *Tonal Pitch Space*. New York: Oxford University Press.

Lerdahl, F. (2004). A Music-Theoretic Approach to the Sounds of Poetry. Ms., Columbia University.

Lerdahl, F., & R. Jackendoff. (1983). *A Generative Theory of Tonal Music*. Cambridge, Mass.: MIT Press.

Nespor, M. & I. Vogel. (1986). *Prosodic Phonology*. Dordrecht: Foris.

Selkirk, E. (1986). On derived domains in sentence phonology. Yearbook of Linguistics **3**: 371-405.

Steriade, D. Knowledge of similarity and narrow lexical override. In P.M. Nowak, C. Yoquelet, & D. Mortensen (eds.), *Proceedings of Berkeley Linguistics Society 29*. Berkeley, CA: BLS. 583-598.